HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P. O. Box 272400
Fort Collins, Colorado 80527-2400

PATENT APPLICATION
Attorney Docket No. 10012041-1

**IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE**

Inventor(s): Jun Gao et al. **Confirmation No.:** 5995
Application No.: 10/050,346 **Examiner:** Yaritza Guadalupe
Filing Date: January 15, 2002 **Group Art Unit:** 2859
Title: CLUSTER-WEIGHTED MODELING FOR MEDIA CLASSIFICATION

**COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450**

DECLARATION OF JUN GAO UNDER 37 CFR 1.132

Sir:

1. My name is Jun Gao. I am an employee of Hewlett-Packard Company.

2. I am a co-inventor of the invention claimed in U.S. Patent Application Serial No. 10/050,346, which is entitled "CLUSTER-WEIGHTED MODELING FOR MEDIA CLASSIFICATION." The other co-inventor of this application is Ross R. Allen.

3. I am also a co-inventor of the invention claimed in U.S. Pat. No. 6,517,180, which is entitled "DOT SENSING, COLOR SENSING AND MEDIA SENSING BY A PRINTER FOR QUALITY CONTROL." The other co-inventors of this issued patent are Ross R. Allen, Barclay J. Tullis and Carl E. Picciotti.

4. It is standard practice at Hewlett-Packard Company to submit an "Invention Disclosure" to the company's legal department when potentially patentable inventions are discovered as a consequence of work performed for the company. Accompanying this Declaration is such an Invention

PAGE 10/26 * RCVD AT 11/18/2003 3:10:10 PM [Eastern Standard Time] * SVR:USPTO-EFXRF-1/1 * DNIS:8729319 * CSID:650 969 6216 * DURATION (mm-ss):16-32

Disclosure. The accompanying Invention Disclosure is labeled Exhibit A and consists of three pages in which entries are made into a standardized form and eleven pages of an explanation of cluster-weighted modeling (CWM) and its application to classifying print media.

5. The invention that is the focus of the Invention Disclosure of Exhibit A is also the focus of the patent application identified in Paragraph 2 of this Declaration.

6. On the first page of Exhibit A, the Invention Disclosure is signed by Ross R. Allen and myself as inventors. There are no other inventors of the disclosed invention.

7. On the second page of Exhibit A, Carl Picciotti, who is a co-inventor of the patent identified in paragraph 3 of this Declaration, is identified as a witness to whom the invention was explained. The "Date of Signature" is February 7, 2001.

8. Each page of the eleven page document that is part of Exhibit A is signed and dated by Ross R. Allen and myself as inventors. The two signatures are dated January 24, 2001.

9. Each page of the eleven page document that is part of Exhibit A is also signed and dated by Carl Picciotti and Raymond Beausoleil. While not apparent from the document, these signatures were witness signatures. The date of the Carl Picciotti signature is February 7, 2001, the same date as that on the second page of Exhibit A for the witness signature of Carl Picciotti.

10. That portion of the issued patent (identified in Paragraph 3 of this Declaration) that describes using information of textural features of a print media as input parameters for a probabilistic input-output model in order to classify the print media is derived from the work of Ross R. Allen and myself

PAGE 11/26 * RCVD AT 11/18/2003 3:10:10 PM [Eastern Standard Time] * SVR:USPTO-EFXRF-1/1 * DNIS:8729319 * CSID:650 969 6216 * DURATION (mm-ss):16-32

and represents the invention described in the Invention Disclosure of Exhibit A.

11. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true and further that the statements are made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: November 18, 2003

Jun Gao

Write in Dark Ink on Front Side Only, Please

INVENTION DISCLOSURE PAGE ONE OF ____

PDNO 10012641 DATE RCVD 2/20/01 ATTORNEY SFH

*Instructions: The information contained in this document is **COMPANY CONFIDENTIAL** and may not be disclosed to others without prior authorization. Submit this disclosure to the HP Legal Department as soon as possible. No patent protection is possible until a patent application is authorized, prepared, and submitted to the Government.*

Descriptive Title of Invention:
Method to Identify Print Media in a Printer

Name of Project:
HPL Labs HardCopy Sensors

Product Name or Number:

Was a description of the invention published, or are you planning to publish? If so, the date(s) and publication(s):
NO

Was a product including the invention announced, offered for sale, sold, or is such activity proposed? If so, the date(s) and location(s):
NO

Was the invention disclosed to anyone outside of HP, or will such disclosure occur? If so, the date(s) and name(s):
NO

If any of the above situations will occur within 3 months, call your IP attorney or the Legal Department now at 1-888-4919 or 970-898-4919.

Was the invention described in a lab book or other record? If so, please identify (lab book #, etc.)

No. See attached document.

Was the invention built or tested? If so, the date:

Yes - July 2000

Was this invention made under a government contract? If so, the agency and contract number:

No

Description of Invention: *Please preserve all records of the invention and attach additional pages for the following. Each additional page should be signed and dated by the inventor(s) and witness(es).*

A. Description of the construction and operation of the invention (include appropriate schematic, block, & timing diagrams; drawings; samples; graphs; flowcharts; computer listings; test results; etc.)
B. Advantages of the invention over what has been done before.
C. Problems solved by the invention.
D. Prior solutions and their disadvantages (if available, attach copies of product literature, technical articles, patents, etc.).

Signature of Inventor(s): Pursuant to my (our) employment agreement, I (we) submit this disclosure on this date: [CIPT(1)].

Employee No.	Name	Signature	Telnet	Mailstop	Entity & Lab Name
00498757	JUN GAO		857-5050	2U-16	0419-5042
00254615	Ross R. Allen		857-8236	2U-19	0419-5042

(If more than four inventors, include additional information on another copy of this form and attach to this document)

Form 3.1 idf.doc, rev. 06/03/00

EXHIBIT A

Write in Dark Ink on Front Side Only, Please

INVENTION DISCLOSURE COMPANY CONFIDENTIAL PAGE ___ OF ___

Signature of Witness(es): *(Please try to obtain the signature of the person(s) to whom invention was first disclosed.)*

The invention was first explained to, and understood by, me (us) on this date: [ ]

Full Name	Signature	Date of Signature
Carl Picciotto	[signature]	2/7/2001
Full Name	Signature	Date of Signature

Inventor & Home Address Information: *(If more than four inventors, include addl. information on a copy of this form & attach to this document)*

Inventor's Full Name: JUN GAO

Street: 437 WHISMAN PARK DRIVE

City: MOUNTAIN VIEW State: CA Zip: 94043

Do you have a Residential P.O. Address? P.O. BOX ___ City ___ State ___ Zip ___

Greeted as *(nickname, middle name, etc.)*: JUN Citizenship: CANADA

Inventor's Full Name: ROSS RICE ALLEN

Street: 408 HAINLINE DR,

City: BELMONT, CA 94002 State ___ Zip ___

Do you have a Residential P.O. Address? P.O. BOX ___ City ___ State ___ Zip ___

Greeted as *(nickname, middle name, etc.)*: ROSS Citizenship: USA

Inventor's Full Name

Street

City State Zip

Do you have a Residential P.O. Address? P.O. BOX City State Zip

Greeted as *(nickname, middle name, etc.)* Citizenship

Inventor's Full Name

Street

City State Zip

Do you have a Residential P.O. Address? P.O. BOX City State Zip

Greeted as *(nickname, middle name, etc.)* Citizenship

Form 3.1 Idf.doc, rev. 06/03/00

Write in Dark Ink on Front Side Only, Please

Description of Invention:	*Please preserve all records of the invention and attach additional pages for the following. Each additional page should be signed and dated by the inventor(s) and witness(es).*

A. Description of the construction and operation of the invention (include appropriate schematic, block, & timing diagrams; drawings; samples; graphs; flowcharts; computer listings; test results; etc.)

See attached document

B. Advantages of the invention over what has been done before.

The present invention provides a highly-reliable, simple, low-cost, and easily embeddable method for identification of the print medium in a printer. It works by microscopic (order of 8 micron pixel) imaging of the surface texture of the print medium. This has been proven to be a reliable method of distinguishing between different types of papers and some overhead transparency films.

Other methods developed for HP DeskJet printers cannot provide the accurate discrimination between similar media types from lower resolution measurements of diffuse and specular reflection.

C. Problems solved by the invention.

HP studies have shown that few users correctly set printer driver dialog box values for the type of paper in the printer. Many never even make this selection. The result is that the print mode of the printer (involving both raster image processing and writing system parameters such as the number of drops of ink per pixel, number of passes, color maps, etc.) is not matched to the paper. The result is inferior print quality. This can lead to significant user dissatisfaction, especially when expensive special media are used.

This invention provides a highly reliable method to match printer settings to the print medium resulting in optimal print quality.

D. Prior solutions and their disadvantages (if available, attach copies of product literature, technical articles, patents, etc.).

Some new models of HP DeskJets (i.e., 990C and PhotoSmart 1220) have a simple media type detector that uses diffuse and specular reflection measured at about 600 dpi (a 40um pixel size) to discriminate between media types. This method is simple and reasonably inexpensive but not as accurate as required in discriminating among different papers.

PAGE 15/26 * RCVD AT 11/18/2003 3:10:10 PM [Eastern Standard Time] * SVR:USPTO-EFXRF-1/1 * DNIS:8729319 * CSID:650 969 6216 * DURATION (mm-ss):16-32

CWM with Application to Media Classification

1. Introduction to Cluster-Weighted Modeling

Cluster-Weighted Modeling is an input/output inference framework based on probability density estimation of a joint set of features and target data. It is similar to mixture-of-experts type architectures and can be interpreted as a flexible and transparent technique to approximate an arbitrary function[1]. During training, clusters automatically "go to where data is" and approximate the subsets of the data space according to a smooth domain of influence. Globally, the influence of different clusters is weighted by Gaussian basis terms, while locally each cluster represents a simple model such as a linear regression function. Thus, previous results from linear system theory, linear time series analysis and other traditional/modern frameworks are applied within the broader context of a globally non-linear model.

Non-linear system modeling uses models with linear coefficients β_m and non-linear basis function $f(x)$,

$$y(x) = \sum_{m=1}^{M} \beta_m f_m(x) \qquad (1)$$

as in, for example, a polynomial expansion. Models may have the coefficients inside the non-linearirities,

$$y(x) = \sum_{m=1}^{M} f_m(x, \beta_m) \qquad (2)$$

as in, for example, a neural network.

In the case of generalized linear model (1), only a single matrix pseudo-inverse is needed to find the set of coefficients yielding the minimum mean-square error. However, the number of coefficients in (1) is exponential in the dimension of x. Equation (2) [2] has more expressive power, which can reduce the number of coefficients needed for a given approximation error to linear in the dimension of x, however, the non-linear parameters of (2) require an iterative search[3].

CWM uses simple local models, which satisfy equation (1), and uses models as described in equation (2) to create global models. Hence, CWM combines the efficient estimation of the former with the benefits of the latter. The local parameters are fitted by a SVD (Singular Values Decomposition) matrix inversion of the local covariance matrix. The remaining cluster parameters in charge of the global weighting are found using a variant of the Expectation-Maximization (EM) Algorithm[4]. EM is an iterative search that maximizes the model likelihood given a data set and initial conditions. A initial starting values for the cluster parameters are picked randomly or according to the application, and then the Expectation step starts.

1/24/01

1/29/01

2/7/01

- ***Expectation-Step:***

In the E-Step, the current cluster parameters are assumed correct and the posterior probabilities that relate each cluster to each data point are evaluated. Those probabilities can be interpreted as the probability that a particular cluster generated a particular data, or as the normalized responsibility of a cluster for a data point:

$$p(c_m \mid y,x) = \frac{p(y,x \mid c_m)p(c_m)}{p(y,x)} = \frac{p(y,x \mid c_m)P(c_m)}{\sum_{l=1}^{M} p(y,x \mid c_l)p(c_l)} \tag{3}$$

where the sum over clusters in the denominator causes clusters to interact, fight over points and specialize in data they best explain.

- ***Maximization-Step***

In the M-step, the current input-output data distribution is assumed correct, and the cluster parameters are found to maximize the likelihood of the data. The new estimate for the unconditioned cluster probabilities is,

$$p(c_m) = \int p(c_m \mid y,x)p(y,x)dydx \approx \frac{1}{N}\sum_{n=1}^{N} p(c_m \mid y_n, x_n) \tag{4}$$

An integral over a density can be approximated by an average over variable drawn from that density. Next, the expected input mean of each cluster is computed, which is the estimate of the new cluster mean:

$$\mu_m = \int x \cdot p(x \mid c_m)dx = \int x \cdot p(y,x \mid c_m)dydx = \int x \frac{p(c_m \mid y,x)}{p(c_m)} p(y,x)dydx$$
$$\approx \frac{1}{N \cdot p(c_m)} \sum_{n=1}^{N} x_n p(c_m \mid y_n, x_n) = \frac{\sum_{n=1}^{N} x_n \cdot p(c_m \mid y_n, x_n)}{\sum_{n=1}^{N} p(c_m \mid y_n, x_n)} \tag{5}$$

The introductions of output y into (5) results that cluster parameters are found with respect to both input and output data. Clusters get pulled based on both where the data is to be explained and how well their model explains the data. For a any function $\theta(x)$, similar to (5),

01/24/01

1/24/01

2/7/01

2/14/01

$$\langle \theta(x) \rangle_m = \int \theta(x) \cdot p)(x \mid c_m) dx \approx \frac{1}{N} \sum_n \theta(x_n) \frac{p(c_m \mid y_n, x_n)}{p(c_m)} = \frac{\sum_n \theta(x_n) p(c_m \mid y_n, x_n)}{\sum_n p(c_m \mid y_n, x_n)} \tag{6}$$

which lead to the cluster weighted covariance matrices,

$$[P_m]_{ij} = \langle (x_i - \mu_i)(x_j - \mu_j) \rangle_m \tag{7}$$

For updating the local model, the model parameters are found by taking the derivative of the *log* of the total likelihood function with respect to the parameters,

$$0 = \frac{\partial}{\partial \beta} \log \prod_{n=1}^{N} p(y_n, x_n) \tag{8}$$

For a single output y and a single coefficient β_m,

$$\begin{aligned} 0 &= \sum_{n=1}^{N} \frac{\partial}{\partial \beta_m} \log p(y_n, x_n) \\ &= \sum_{n=1}^{N} \frac{p(y_n, x_n, c_m)}{p(y_n, x_n)} \cdot \frac{y_n - f(x_n, \beta_m)}{\sigma^2_{m,y}} \cdot \frac{\partial f(x_n, \beta_m)}{\partial \beta_m} \\ &= \frac{1}{N \cdot p(c_m)} \sum_{n=1}^{N} p(c_m \mid y_n, x_n)[y_n - f(x_n, \beta_m)] \frac{\partial f(x_n, \beta_m)}{\partial \beta_m} \\ &= \left\langle \left[y - f(x, \beta_m) \frac{\partial f(x, \beta_m)}{\partial \beta_m} \right] \right\rangle_m \end{aligned} \tag{9}$$

Put equation (1) into (9), the expression to update β_m is obtained,

$$\begin{aligned} 0 &= \langle [y - f(x, \beta_m)] f_j(x) \rangle_m \\ &= \underbrace{\langle y f_j(x) \rangle_m}_{a_{j,m}} - \sum_{i=1}^{M} \beta_{m,i} \underbrace{\langle f_j(x) \cdot f_i(x) \rangle_m}_{B_{j,i,m}} \end{aligned} \tag{10}$$

01/24/01

1/24/01

2/14/01

2/7/01

$$\Rightarrow \beta_m = \mathrm{B}_m^{-1} \cdot a_m \qquad (11)$$

For a whole set of model parameters, equation (11) expands to,

$$\beta_m = \mathrm{B}_m^{-1} \cdot \mathrm{A}_m \qquad (12)$$

with,

$$[\mathrm{B}_m]_{ij} = \langle f_i(x,\beta_m) \cdot f_j(x,\beta_m) \rangle_m \qquad [\mathrm{A}_m]_j = \langle y_j \cdot f_j(x,\beta_m) \rangle_m \qquad (13)$$

and finally, the output covariance matrices associated with each model are estimated by,

$$\mathrm{P}_{y,m} = \left\langle [y - \langle y|x \rangle]^2 \right\rangle_m = \left\langle [y - f(x,\beta_m)] \cdot [y - f(x,\beta_m)]^T \right\rangle_m \qquad (14)$$

- ***Summarize CWM Algorithm Model Estimation Process (E-M Iteration)***

1. Pick some initial conditions and initial cluster value;
2. Evaluate the probability of the data $p(y,x|c_m)$;
3. Find the posterior probability of the clusters $p(c_m|y,x)$;
4. Update
   (i) the cluster weights $p(c_m)$;
   (ii) the cluster-weighted expectations for the input means μ_m^{new};
   (iii) variance $\sigma_{m,d}^{2,new}$, or covariance P_m^{new};
   (vi) the maximum likelihood model parameters β_m^{new}, and finally
   (v) the output variances $\sigma_{m,y}^{2,new}$
5. Go back to 2, until the total data likelihood does not increase anymore.

2. The CWM Algorithm in a Practical Media Identification Sensor

(1) A Practical Media Identification Sensor

A practical media identification sensor allows a printer to determine the type of paper (i.e., "print medium") in the print zone or paper tray and to adjust the print engine parameters accordingly for optimal print quality. Furthermore, identification of the presence of certain types of transparency film or special papers can be used to prevent damage to the print engine. For example, the coatings on some ink jet transparency films can melt on the fuser roller of a electrophotographic (e.g., HP "LaserJet") printer causing damage that requires the fuser roller to be replaced.

Easily observed using a microscope and grazing illumination (e.g., 45° to 75° from the surface normal) is the surface texture of papers and some transparency films. This surface texture has

---

[1] The matrix inversion should be done by SVD (Singular Value Decomposition) to avoid possible numerical problem with singular covariance matrices

01/24/01 01/24/01 02/14/01 2/7/01

features with characteristic sizes ranging between about 5 μm to about 100 μm. Each type of print medium has a characteristic surface texture, and that provides the fundamental principle for media identification using analysis of microscopic surface features.

An automatic device to discriminate among and thereby identify print media can be built using an image sensor employing a single pixel, a line of pixels, or a two-dimensional array of pixels. Depending upon the size of the sensor's pixel(s), optics image a specified area on the medium's surface onto the pixel. Typically, the viewed area of the print medium surface is a square about 5 μm to about 100 μm on a side with 10-40 μm giving practical results.

Surface texture can be characterized by a collection of measured gray-level values obtained by multiple samples over an unprinted area of the print medium's surface. Multiple samples may be obtained by scanning a single pixel sensor over the medium surface and taking measurements at different places (cf. patent applications by Steve Walker, HP VCD) or with a linear or area array of pixels. An advantage of a line or area sensor over a single pixel sensor (e.g., a photodiode, a phototransistor, or integrated "light-to-voltage" or "light-to-frequency" sensor) is multiple samples over a region of the print medium's surface may be obtained without requiring relative motion between the sensor and the print medium. This is useful for simplifying the mechanism for identifying the print medium in the input tray, where no motion source is generally available until the medium is fed into the paper path. Alternatively, a single pixel, line, or area sensor may accumulate multiple samples of the surface of the print medium as the print medium is fed from the input tray into the paper path, or at a point along the paper path where the medium passes under the (fixed) sensor, or by placing the sensor on a scanning print carriage where it moves across the stationary print medium. The objective of all these implementations is to accumulate multiple samples at different locations so as to evaluate variation in surface texture. In general, the objective is to improve the sampling statistics with more samples.

The image sensor preferably has its optical axis along the normal to the plane of the print medium and captures an image of the surface illuminated by multiple wavelengths, for example those produced by green and blue LEDs, arranged to provide illumination at an incidence angle between about 75^0 and 45^0 from the surface normal. These LEDs are illuminated sequentially and pixel measurements are taken under each illuminant. More accurate identification involves the use of multiple illumination sources at different incidence angles[1]. Practically, this sample can be made over the identical physical region or over a different region.

A practical method uses one or more (LED) illuminants, a linear array of 80 pixels sampling at a single fixed location on the print medium surface, and optics to image a region of ~8um square onto each pixel. This arrangement has provided reliable results without the necessity of moving the medium relative to the sensor or vice versa.

The mean of the gray-level values of pixel data and their standard deviation are taken from images of microscopic surface features under illuminants with different wavelengths and angles of incidence. The mean value is the average reflectivity of the media and the standard deviation represents a measure of the texture roughness of the media. Training is required to establish a

[1] A development version of the prototype media identification sensor used four LED illuminators: Green, Blue at a 45^0 angle and Red, Infrared at a 75^0 angle with respect to the surface normal. This provided more experimental flexibility during development and more degrees-of-freedom in the sampled data.

01/24/01

LUT (look-up-table) for different media types/groups[1], and when an unknown media is sampled, the media ID sensor identifies the sample by finding its closet match to the reference set.

(2) CWM Algorithm for a Practical Media Identification Sensor

A group of different media types for LaserJet or Inkjet printer can be characterized as a non-linear, non-Gaussian, complex multi-dimension input-output system. The input of the system is the mean/standard deviation (μ/σ) pair computed from the pixel data sampled by the sensor from two different illuminants at specified angles of incidence. The output is the best match of the unknown print medium to reference media types or media groups.

One of the easier ways to achieve the goal of media identification is to take enough data samples from each media type and compute the means and standard deviations for each illuminant at its angle of incidence. Then, the mean and standard deviation of the means and standard deviations for each media type is computed and stored in the LUT. When a new set of (μ/σ) data from two illuminants are computed from an unknown media type, the distances of the (μ/σ) of the new data to those of the media types in the LUT are calculated, and the media type/group is then determined by some function of these distances: the simplest solution is to find the minimum distance. This approach is similar to using the same numbers of clusters as the numbers of media types in CWM algorithm. This approach provides satisfactory results only if the media data clouds are relatively symmetric and non-singular. Otherwise, the error of this approach can be big and further reduction of error is impossible. As shown in figure 1, which is the (μ/σ) data clouds for seven LaserJet media types, the assumption that each media type is symmetric and non-singular in (μ/σ) domain is not accurate. In fact, it is just the opposite. The data cloud is very asymmetric and singular.

The CWM framework provides an idea solution for this problem. In CWM, the input vector x_i is defined for samples taken under green and blue illuminants as:

$$x_i = \left[\mu_{green}\ \sigma_{green}\ \ \mu_{blue}\ \ \sigma_{blue}\right]_T$$

and the output vector, in this case, a scalar y is one of the media types. In training, the set of vector pairs $\{y_i,\ x_i\}_{i=1}^{N}$ are used to train the CWM input-output model, using a simple local model, $y = \beta_m \cdot x$. In the synthesis process, an unknown input vector x_j is applied to the predictor, which calculates $p(y, x_j)$ according to the trained CWM model to provide the probabilities of that input vector with respect to all media types in the training set. The probabilities of the unknown media with respect to different media group can be further summed up by adding all probabilities for the media types that belong to the media group.

The training process is both time-consuming and computational intensive, especially in the process of gathering all different media samples. It takes several thousand input vectors for each media type to provides a good estimate of the media distribution (i.e., "the data cloud"). It is computational intensive because of the required statistics calculations and matrix manipulations.

[1] *Media Type* refers to an individual LaserJet or Inkjet media product, for example, HP Premium Glossy Photo Paper, HP Multipurpose Paper, Xerox Xpression Paper, etc. *Media Group* refers to media types that have similar recording characteristics and would use similar print engine parameters such as drop volume, number of drops per pixel, etc.

1/24/01 01/24/01 2/7/01

Fortunately, this process can be carried out off-line and only *once for all media types/groups to* be used for a particular printer. The training process is updated only when new media types/groups are introduced, or with changes in the optic/electronics design of the media sensor.

It is practical to train a printer to new media types if bidirectional communications exist between a printer and its host computer and appropriate software is installed on the host. In this case, the training for additional media types could occur during a time when the printer is idle. The media identification sensor would provide the raw pixel data to the host for processing and association with the new media type sample. Software to accomplish this task could be conveniently downloaded from the Internet or be shipped with the printer as part of a printing solutions software application.

In any case, after training a small predictor (on the order of a few 100 lines of C code) and the ensemble of cluster parameters are all that is needed to implement an embedded media identification solution. This entire process could execute within the printer or on the host. In this case, the printer resources must include some image processing capability to optimize the raster image data for rendering by a particular print algorithm (e.g., for ink jet: dot levels per pixel, number of print carriage passes; for electrophotographic printers: feed rate, fuser temperature, etc.). In this case, partially-rendered image data is presented to the printer by the host and the printer completes the rendering process. This method is used in some HP ink jet printers (e.g., DeskJet 980, 990, 2200) with the so-called "HP High Performance Architecture." In another method, the pixel measurements are uploaded to the host from the printer for processing. The raster image processing is done in the host and a fully-rendered image is sent to the printer. In principle, some combination of these processes could be used.

The size of the cluster parameters is determined by the dimensions of input and output. Therefore, the LUT for CWM is determined by numbers of clusters used and the dimensions of input-output vector pair. The LUT should be relatively small - a few Kbytes. Therefore, the whole CWM implementation in printer of media sensor should have a footprint of several Kbytes, which is extremely small by current memory standards. It also should be relatively fast therefore minimizing impact on throughput.

(3) Preferred Cluster-Weighted Modeling Implementation on the Prototype Media Sensor

1. **Sensor Optic Focus Calibration**: The optics are designed and focused to ensure that the pixel resolution of 8 μm square on the medium surface with an optical blur circle of about 20 to 25 μm can be achieved.

2. **Sensor Calibration**: There are several noise sources in any image sensor and the data acquisition system, which must be eliminated or reduced as much as possible. The major source of noise are (1) sensor electronic noise (dark current); (2) sensor photon shot noise; (3) pixel-to-pixel variation; and (4) illumination non-uniformity caused by the source. The first two noise sources are random in nature and can only be effectively reduced by averaging. Their impact to the measurement is minor with the choice of adequate illumination levels. Sensor pixel-to-pixel noise is a fixed, high spatial frequency noise, and the illumination non-uniformity is fixed low spatial frequency effect. These two noises are significant and must be addressed. The method to reduce these effects involves taking samples from imaging a white tile illuminated at several intensity levels.

2/14/01 1/24/01 1/24/01 2/7/01

The high- and low frequency effects are separated and a correction LUT (with values depending on average illumination) is applied to individual pixel outputs.

3. **Black Backing for the Measurement:** A black tile is required to back up each sheet of print medium sample during measurement. This eliminates effects of light penetrating multiple sheets and provides a consistent and optimized sampling environment. It is important that the optical absorption characteristics of the tile used in training be identical to that used in the practical measurement. The black tile could be conveniently replaced with an opening leading into a nonreflective chamber, which should provide similar result.

4. **Training Set Generation:** When training, an area sensor imager can be used to speed up the sampling process. The area imager sensor data is then sub-sampled to the same numbers of pixel as the final production line sensor. An input vector for blue & grean illumination, for example, $\{y_{media\_type}, \mu_{green}, \sigma_{green}, \mu_{blue}, \sigma_{blue}\}$ is then computed from the sub-sampled pixels. For each media type, a few thousands of these input vectors are required to reconstruct reliably the data clouds of that particular media type[1]. Once all different media types are sampled, the input vectors are randomized to provide a better training set.

5. **Optimization of CWM Algorithm for Media Classification:**
   a. ***Initial Location of the Clusters***: Clusters should not be initialized arbitrarily since the algorithm only guarantees to terminate in a local likelihood maximum. The clusters should be placed as close to their final position as possible to save training time and get better convergence of data. The method of selecting initial cluster position is as follow: Choose 1/N as the initial cluster probabilities, where N is the numbers of the clusters. Pick randomly as many points from the training set as there are clusters and initialize the cluster input mean, as well as the cluster output mean with these points. Set the remaining output coefficients to zero. Use the size of the data set in each space dimension as the initial cluster variances.
   b. ***Normalization***: It is required to normalize the training set to zero main and unit variance since arbitrary data values may cause probabilities to because too small.
   c. ***Optimize Numbers of Clusters and Numbers of EM Iteration***: There is no rule as to how many clusters is optimal to a specific problem. Numbers of clusters should be larger than numbers of distinguishable output, in this case, numbers of media types. However, more clusters do not mean better discrimination. With too many small clusters, establishing membership may be difficult especially when a region is populated with many small clusters belonging to different media types. The same can be said for numbers of training iterations between EM steps when the number of cluster is constant. Therefore, an iterative search of increasing numbers of clusters from small to large and numbers of training iterations from small to larger for each particular numbers of clusters has to be performed and determined empirically. For example, with a sample of 7 LaserJet media, it was determined that 24 clusters and 23 iterations were optimal, and this provided the

[1] In our testing, we took 4800 input vectors for each media types. The area sensor we used is 80x80 pixel and the sampled data is then sub-sampled 80 times to generate 80 80-pixels data points, which in turn generated 80 input vectors.

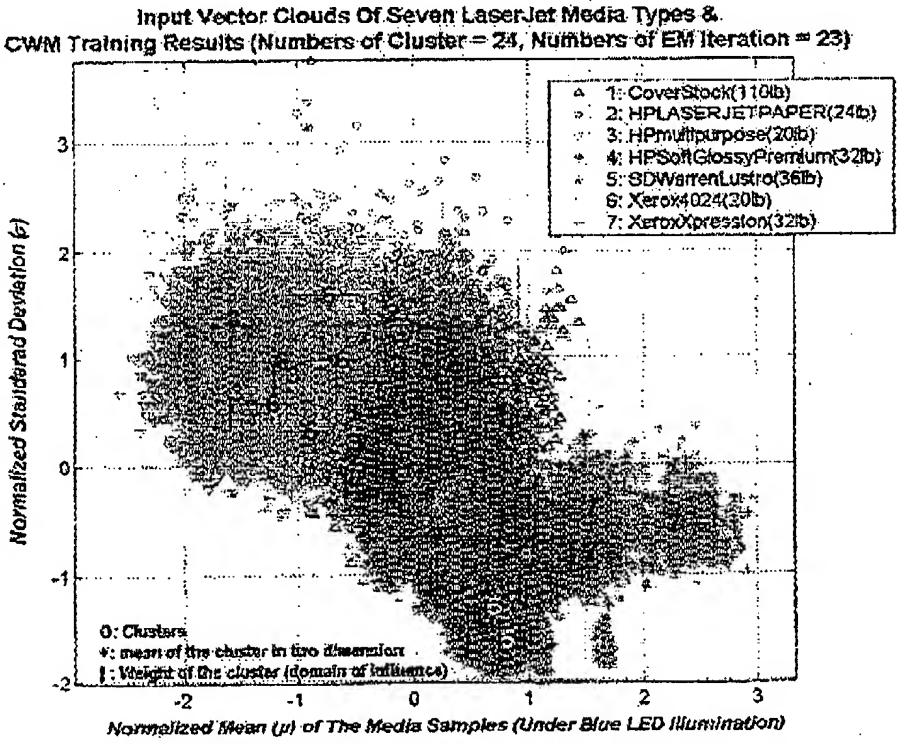
1/24/01 2/14/01 01/24/01 2/7/01

highest correct classification rate (refer to figure 1). For 4 inkjet media, 6 clusters and 10 iterations provided the ultimate results of 100% correct media identification (refer to figure 3).

(4) Media Classification Using CWM Algorithm - Practical Results

There were two cases of media classification studied by HP Labs for LaserJet media and Inkjet media. It was obvious that inkjet media shows a much better separation than the LaserJet media (refer to figure 1 and 3).

Out of 7 LaserJet media types, there are two media groups, namely HP Multipurpose plain paper (less than 30 lb, except CoverStock paper, which is 110lb) and high quality glossy paper (all over 30 lb). The cluster distribution and the final training and testing results for this case are shown in figure 1 and figure 2. It is obvious from figure 1 that, at least for the partial input vector $\{v \quad \mu_{blue} \quad \sigma_{blue}\}$, the media samples are all mixed and the data clouds have no clear boundary. The overlapping and mixing of different media types are not nearly as bad as shown in figure 1 when the whole input vector (including the additional two dimensions for green illumination) is considered. But still, it is a difficult system for estimation and prediction. With 24 clusters to estimate 7 different media types, after 23 EM iterations, the clusters settled down to the data clouds shown and it is clear that there are more than one cluster for each media types.



Input Data Clouds Computed From Seven LaserJet Media Samples & Cluster-Weighted Estimation Results

1/24/01

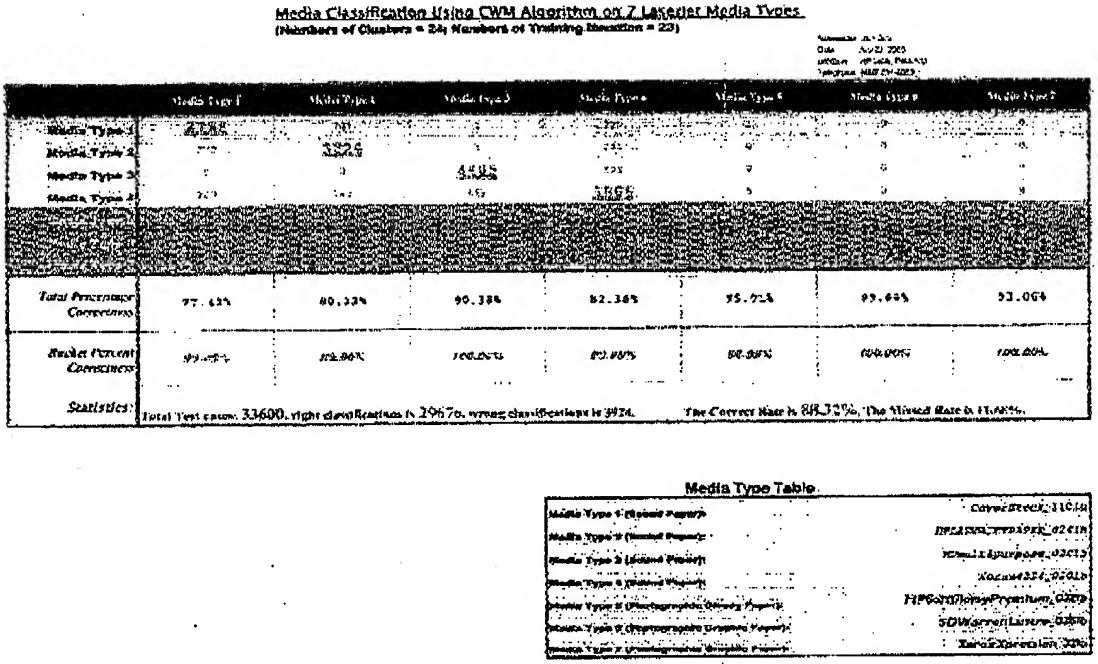
2/17/01

01/24/01

2/17/01

Figure 2 shows the training and testing results for the 7 media types. The yellow rows are the media types in the *multipurpose plain paper group* (Types 1-4) and the gray rows are the three media types in *high quality glossy media group* (Types 5-7). The rate for correctly identifying an individual sample ranged from 77% to 99%. There was very high accuracy of correctly identifying membership of a sample in either the *multipurpose* group or the *glossy* group.

It is also possible to use CWM algorithm and the microscopic sensor to determine roughly the weight of the media in the printer tray. This is obtained by associating the weight of the print media sample with the surface texture characteristics by which it will be identified. This is an important information for LaserJet print engine since the printing mode and speed is determined, ideally by the weight of the media.



CWM for Media Classification Testing Results for Seven LaserJet Media Types in Two Media Groups

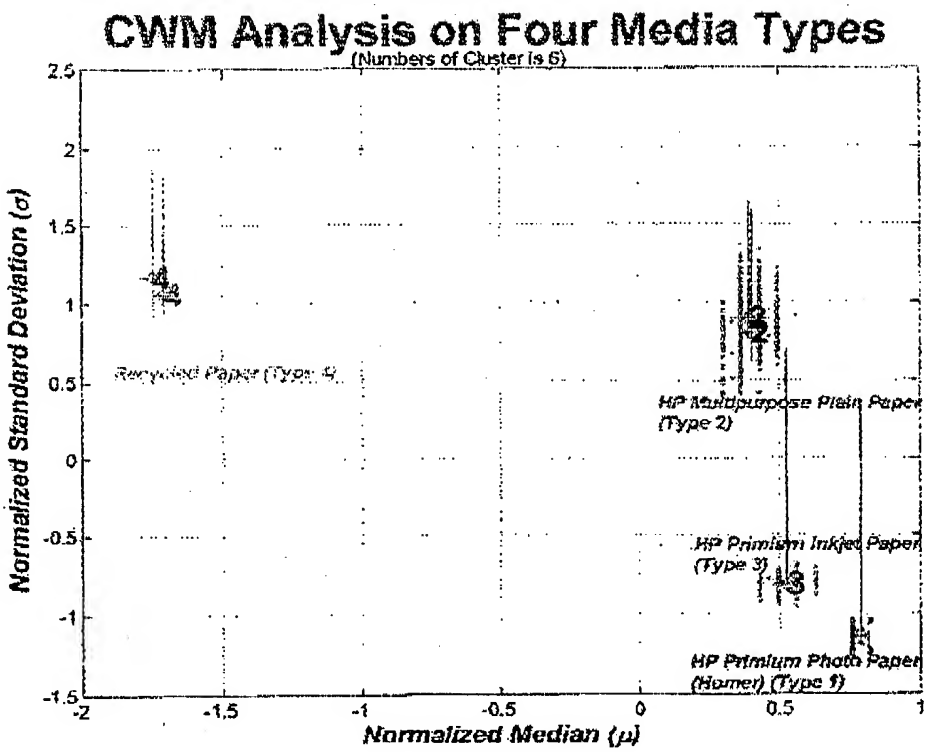
The 4 Inkjet media case is shown in figure 3 and is included to demonstrate the effect of media surface characteristics on cluster geometry. The result is that the identification problem is easier since there is less aliasing between clusters compared to the LaserJet media tested: the data clouds are significantly separated, even in this partial input vector space; and a much less numbers of cluster is required to fully estimate and predict the membership of a sample among the 4 different types and groups of inkj t media studied.

1/24/01

2/14/01

01/24/01

2/7/01



Input Data Clouds From Four Different Inkjet Media Types & CWM Estimation Results

---

[1] B. Schoner, C. Cooper, C. Douglas, N. Gershenfeld. *Data driven Modeling and Synthesis of Acoustical Instruments*. Proceedings of the International Computer Music Conference, Ann Arbor, Michigan, 1998

[2] Andrew R. Barron. *Universal Approximation Bounds For Superposition Of A Sigmoidal Function*. IEEE Transaction On Information Theory, 39:930-945, 1993

[3] Neil Gershenfeld. *The Nature Of Mathematical Modeling*. Cambridge University Press, New York, 1999

[4] A. P. Dempster, N. M. Laird and D. B. Rubin. *Maximum Likelihood From Incomplete data via the EM Algorithm*, J. R. Statist. Soc. B, 39:1-38, 1977

1/24/01 01/24/01

2/14/01 2/7/01